

MOLECULAR SYSTEMATICS OF *CLERODENDRUM* (LAMIACEAE): ITS SEQUENCES AND TOTAL EVIDENCE¹

DOROTHY A. STEANE,^{2,4} ROBERT W. SCOTLAND,²
DAVID J. MABBERLEY,^{2,5} AND R. G. OLMSTEAD^{3,6}

²Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK;

³Department of Environmental, Populational and Organismic Biology, University of Colorado, Boulder, Colorado 80309

Thirty-three species of *Clerodendrum* s.l. and five outgroup genera were included in a sequence analysis of internal transcribed spacers of the nuclear ribosomal DNA. The results of the cladistic analysis were compared to and combined with cpDNA restriction site data from a previous study. All molecular data identified four major clades within *Clerodendrum* s.l. and showed the genus to be polyphyletic. *Clerodendrum* s.s., minus *Konocalyx* and *Cyclonema*, is monophyletic and the genus should be restricted to this group. *Cyclonema* and *Konocalyx* form a clade distinct from *Clerodendrum* s.s., which has been recognized as *Rotheca* Raf.

Key words: chloroplast DNA; *Clerodendrum*; *Cyclonema*; internal transcribed spacer; Lamiaceae.

Clerodendrum L. (Teucrioideae, Lamiaceae, sensu Cantino, Harley, and Wagstaff, 1992) is a pantropical/subtropical genus, predominantly distributed in Africa, Asia, and Pacific Oceania, with fewer representatives in the New World. The genus, comprising 400–500 specific and subspecific taxa, displays a high degree of morphological variation. Most classifications distinguish two large subdivisions within the genus, and as many as four smaller groups, depending on the author (e.g., Schauer, 1847; Briquet, 1897; Lam, 1919; Thomas, 1936). The two larger groups, *Clerodendrum* and *Cyclonema* (Hochst.) Gürke, have been ranked at sectional (Briquet, 1897), subgeneric (Thomas, 1936), and generic (Schauer, 1847; Cantino, Harley, and Wagstaff, 1992) levels. Phenetic and cladistic studies have led to the suggestion that *Clerodendrum* s.l. (sensu lato) is paraphyletic (Stenzel et al., 1988; Cantino, 1992) or polyphyletic (Winterhalter,

1991; Cantino, 1992; Rimpler, Winterhalter, and Falk, 1992), although until recently (Steane et al., 1997), conclusive evidence for this was lacking.

Two studies of cpDNA restriction sites (Steane et al., 1997) and sequences (Wagstaff et al., 1998) showed *Clerodendrum* s.l. to be polyphyletic. These results indicated that subgenus *Cyclonema* (Hochst.) Thomas and section *Konocalyx* Verdc. (subgenus *Clerodendrum* pro parte) formed a monophyletic group, emerging as a lineage distinct from the rest of *Clerodendrum*. Sequence data from *ndhF* indicated that *Tetraclea* is nested within *Clerodendrum* s.s. (sensu stricto) (i.e., subgenus *Clerodendrum* excluding section *Konocalyx*), therefore indicating that *Clerodendrum* s.s. is paraphyletic (Steane et al., 1997).

There is still uncertainty, however, surrounding the putative polyphyly of *Clerodendrum* s.l. The morphological studies of Cantino (1992) and Rimpler, Winterhalter, and Falk (1992; this study also included phytochemical data) both found *Clerodendrum* to be non-monophyletic, but there is substantial discord between the findings of the two studies with respect to relationships within *Clerodendrum* s.l. and between *Clerodendrum* s.l. and other labiate genera. Furthermore, the strict consensus of Cantino's (1992) study was too poorly resolved to determine whether *Clerodendrum* is paraphyletic or polyphyletic. Sampling of *Clerodendrum* s.l. in the sequencing study of Wagstaff et al. (1998) was restricted to two species, and more comprehensive sampling is required to confirm the polyphyly of *Clerodendrum* s.l. Finally, one criticism of the cpDNA-based phylogenies centers on the uniparental mode of chloroplast inheritance (Harris and Ingram, 1991; Mogensen, 1996). Phylogenetic reconstructions based on cytoplasmically inherited genomes may be susceptible to significant error from hybridization and introgression or lineage sorting (Rieseberg and Soltis, 1991; Doyle, 1992). Comparisons of trees derived from nuclear DNA with those derived from cpDNA may assist in the identification of such problems, as well as providing corroborative evidence of relationships hypothesized by cpDNA data.

Steane et al. (1997) identified four distinct groups

¹ Manuscript received 18 December 1997; revision accepted 22 May 1998.

The authors thank those gardens and individuals who supplied plant material for this research: Joseph Reynold O'Neill Botanic Garden, British Virgin Islands; Townsville Botanic Garden, Australia; John Grimshaw, Kilimanjaro Elephant Project, Tanzania; University of California Botanic Garden, Berkeley, California; D. Willman, Institut für Pharmazeutische Biologie, Albert-Ludwigs-Universität, Freiburg, Germany; Isidro E. Meúdez, National Botanic Garden, Cuba; Ann Robertson, Malindi, Kenya, for providing plant specimens for DNA extraction. We would like to thank D. Swofford (Smithsonian Institution) for providing a test version of PAUP* and Phil Cantino (Ohio University, Athens) for helpful discussions. This research was funded by: Rhodes Trust, Oxford; Trapnell Fund, Oxford; The Committee for Graduate Studies, University of Oxford; The Nancy Lindsay Trust Fund, Oxford; The Percy Sladen Memorial Fund; University College, Oxford; United States National Science Foundation grant number BSR9107827 (to R.G.O.); and SGER grant number DEB9224126 to Fairchild Tropical Garden, Florida.

⁴ Author for correspondence, current address: Cooperative Research Centre for Sustainable Production Forestry, School of Plant Science, University of Tasmania, G.P.O. Box 252-55, Hobart, Tasmania 7001, Australia.

⁵ Current address: Royal Botanic Gardens Sydney, Mrs. Macquarie's Road, Sydney, 2000, Australia.

⁶ Current address: Department of Botany, Box 355325, University of Washington, Seattle, WA 98195-5325.

within *Clerodendrum* s.l., three of which together form a clade. The fourth forms a monophyletic group distinct from the other three and probably deserves delimitation as a separate genus. These findings require corroboration before evolutionary or biogeographic interpretations can be made.

The internal transcribed spacers (ITS) of the nuclear ribosomal DNA repeat (nrDNA) are two regions of non-coding and relatively rapidly evolving DNA sequence that flank the very slowly evolving 5.8S ribosomal RNA gene. The region comprising the ITS and 5.8S gene has been used extensively for phylogenetic inference among relatively closely related species (e.g., Gonzalez et al., 1990; Lee and Taylor, 1992; Baldwin, 1992, 1993; Baldwin et al., 1995; Suh et al., 1993; Wojciechowski et al., 1993; Yuan, Küpfer, and Doyle, 1996).

The nrDNA sequence data presented in this paper are compared to and combined with restriction site data from the chloroplast genome presented in a previous study (Steane et al., 1997). There are two basic approaches to combining independent data sets, or “process partitions” (i.e., subsets of characters with different evolutionary and biological properties; Bull et al., 1993; Miyamoto and Fitch, 1995), for the same group of organisms. Firstly, sets of trees from separate analyses of different data sets may be combined and a consensus tree computed. Such consensus methods provide an indication of the congruence among trees produced from the different data sets and may be useful for comparing competing hypotheses of relationship, e.g., those arising from organelle-based analyses and those based on other data sets (Miyamoto, 1985; Hillis, 1987; Miyamoto and Fitch, 1995). Miyamoto and Fitch (1995) argue that the evolutionary and biological properties of different data sets make it likely that agreement among their resulting topologies is the result of the true species phylogeny rather than of similar, but nonphylogenetic, factors responsible for the separate histories of character sets, or of systematic errors and model failures in phylogeny reconstruction methods. Hence, they emphasise corroboration between independent data sets as a means to formulate phylogenetic hypotheses.

Alternatively, independent data sets may be combined and analyzed simultaneously. This provides an assessment of the overall congruence of characters from all sources of data and may enhance detection of the true phylogeny. By increasing the number of characters in an analysis, phylogenetic signal may assert itself over the noise (coincidental similarity due to homoplasy) from each individual data set, resulting in a more accurate estimate of true phylogeny (Barrett, Donoghue, and Sober, 1991; Mishler, 1994; Olmstead and Sweere, 1994). Combined analysis may resolve conflict among trees from separate analyses (e.g., Kim and Jansen, 1994), may resolve phylogeny at different levels (e.g., Pennington, 1996), and may reveal groups not present in any of the separate trees. However, if there is heterogeneity among data sets with respect to some property that affects phylogeny estimation (e.g., differences in substitution rate; Bull et al., 1993), then combining the data can give misleading results (de Queiroz, Donoghue, and Kim, 1995). Where different data sets yield strongly supported but conflicting cladograms it may be judicious to keep the

data sets separate. The Incongruence Length Difference (ILD) test of Farris et al. (1994) assesses the heterogeneity of data sets and gives an indication of whether there is good reason for keeping them separate.

The goals of this study were to: (1) use nuclear DNA to corroborate results derived from chloroplast DNA (Steane et al., 1997); (2) use consensus methods to identify areas of agreement between the cpDNA and the nrDNA of *Clerodendrum*; and (3) use congruence methods by combining data from both sources to provide the most accurate reflection of *Clerodendrum* phylogeny based on all available molecular evidence.

MATERIALS AND METHODS

Thirty-three of the 79 accessions of *Clerodendrum* that had been included in the cpDNA restriction site analysis (Steane et al., 1997) were selected for inclusion in a sequence analysis of the 5.8S nrDNA and the flanking ITS regions. All subgenera and sections of *Clerodendrum* s.l. were represented except for subgenera *Kalaharia* and *Tridens*, section *Eurycalyx* (Thomas) Verdc. (subgenus *Clerodendrum* pro parte), and sections *Racemiflora*, *Oligocymosa* Thomas, and *Pleurocymosa* Thomas (subgenus *Cyclonema* pro parte). Putative outgroups for the analysis were selected on the basis of existing theories of relationships (Cantino, 1992) and the relationships inferred by cpDNA (Steane et al., 1997), although sampling was constrained by the availability of material. *Karomia* Dop. (Clerodendreae), *Trichostema* L., *Caryopteris* Bunge (Caryopterideae), *Oxera*, and *Faradaya* (Clerodendreae) were included as outgroups.

The DNA samples used in the sequence analysis were the same as those used in the cpDNA analysis of Steane et al. (1997). The 5.8S nrDNA and flanking ITS regions were amplified using the polymerase chain reaction (PCR) with primers ITS 5 and ITS 4 (White et al., 1990). The amplification conditions were those described by Baldwin (1992): 97°C for 1 min, 48°C for 1 min, and 72°C for 45 s increasing by 4 s/cycle over 40 cycles. Single-stranded DNA was produced by including 10 µL of double-stranded DNA in a second 100-µL reaction mixture containing only one of the two primers (Kaltenboeck et al., 1992). Twenty-five cycles of PCR were required for the single-stranded amplifications.

Single-stranded PCR products were sequenced with TAQuence (Amersham, Arlington Heights, Illinois), using ³²P dATP, in accordance with the recommendations of the manufacturer. The sequencing reactions were primed using the ITS 5 and ITS 4 primers externally, and ITS 3 and ITS 2 internally. Both strands of DNA were sequenced. To overcome band compressions in the gels, reactions containing 7-deaza-dGTP were run in addition to reactions containing dGTP.

Sequences were aligned using the DNA sequence alignment program Clustal V (Higgins, Bleasby, and Fuchs, 1991), followed by visual inspection (GenBank database accession numbers GBANU77739–GBANU77768). Taxa included in the study have highly divergent ITS sequences, with numerous substitutions, insertions, and/or deletions. Two sets of similar sequences emerged, one of which comprised taxa from subgenus *Cyclonema* and section *Konocalyx* (subgenus *Clerodendrum* pro parte). The other group comprised the remaining *Clerodendrum* taxa. The sequences within each set were of similar length and could be aligned easily. One particular sequence tract (~120 bp long) was aligned easily within each group of taxa but not between the groups. In the alignment used in the phylogenetic analyses, a corresponding series of gaps was inserted in each alternate group of taxa in this region. This allowed the use of all reliable data to resolve relationships within each of the groups, as well as using the other data that were alignable between the groups for higher level resolution. Sequence within this region was omitted from the outgroup taxa, because alignment with either one of the two groups of *Clerodendrum* was ambiguous.

TABLE 1. Numbers of informative and uninformative characters included in the combined data set.

Characters	Number of characters		
	cpDNA data	nrDNA data	Total data
Informative	329	285	614
Uninformative	215	623	838
Constant	24	431	455
Autapomorphic	191	192	383
Total	544	908	1452
Missing data ^a	3.1%	7.1%	5.6%

^a "Missing data" refer to the percentage of cells in the data matrix scored as "missing" (does not count gaps in the sequence alignment).

Pairwise percentage sequence divergence values were calculated by dividing the patristic distance between the taxa (based on the alignment between all taxa) by the total number of base pairs (bp) sampled in each data set (i.e., 913 bp in the nrDNA sequence data set and 4890 bp in the cpDNA restriction site data set) and multiplying by 100.

Chloroplast DNA restriction site data for the 33 accessions used in the present study were obtained from an earlier study (Steane et al., 1997). These data were analyzed independently of and in conjunction with the nrDNA sequence data obtained in the present study. Data sets from this smaller cpDNA analysis and the nrDNA sequence analysis were combined, resulting in a data matrix comprising 38 taxa and 1452 characters (Table 1).

Parsimony analyses were undertaken using PAUP vers. 3.1.1 (Swofford, 1991). Heuristic searches were carried out using 100 replicates of a random addition sequence, followed by TBR (tree bisection and reconnection) branch swapping with steepest descent (MULPARS on). A strict consensus tree was computed from each set of equally most parsimonious trees. Each data set was bootstrapped 250 times, using closest addition sequence, followed by TBR branch swapping with steepest descent (MULPARS off). The consensus analysis was carried out by combining the sets of trees that had been obtained in previous analyses and calculating a semistrict consensus. The ILD test of Farris et al. (1994; as implemented in PAUP* v. 4.0d56, kindly provided by D. L. Swofford) was used to test for incongruence between the data matrices.

The data matrices are available from treeBASE [study accession number 227; matrix accession number M248 (ITS); and matrix accession number M249 (cpDNA)] and the first author.

RESULTS

The boundaries of the internal transcribed spacers and the nrDNA coding regions in the 38 accessions were identified by comparison with those of tomato [*Solanum lycopersicum* L. (Solanaceae); Kiss et al., 1988]. For most taxa, ~650 nucleotides were sequenced; the aligned sequence length, including an ~120-bp gap (see Materials and Methods), was 913 bp. The length of the 5.8S nrDNA was 165 bp in *Faradaya*, *Oxera*, and all *Clerodendrum* taxa for which data were available, except *C. acerbianum*, which was 166 bp long. These lengths are close to those reported for other taxa [e.g., 163 in tomato (Kiss et al., 1988); 164 bp in members of Winteraceae (Suh et al., 1993) and *Calycadenia* (Compositae; Baldwin, 1993)]. The 5.8S nrDNA of *Caryopteris*, *Karomia*, and *Trichostema* were not sequenced completely. The length of the ITS 1 region in the taxa surveyed ranged from 200 bp in *C. paniculatum* to 230 bp in *C. glabrum* (Table 2). Within *Clerodendrum* ITS 2 ranged from 212 bp in *C. incisum* to 219 bp in *C. makanjanum* and *C. myricoides* ssp. *muenzneri* (Table 2). The longest ITS 2

TABLE 2. Approximate sizes of internal transcribed spacers.

Clade/taxon	Size range (bp)	
	ITS 1	ITS 2
Clade I	200–203	217–222
Clade II	211–216	218–224
Clade III	211–230	217–223
Clade IV	218–229	212–219
<i>Faradaya</i>	207	216
<i>Oxera</i>	? ^a	214
<i>Caryopteris</i>	? ^a	235

^a Question marks indicate sequence not complete for region indicated.

sequence was in *Caryopteris* (235 bp). The size ranges of both ITS 1 and ITS 2 for *Clerodendrum* are within the ranges reported for other taxa (Baldwin, 1993; Baldwin et al., 1995; Suh et al., 1993; Wojciechowski et al., 1993).

Estimated sequence divergence ranged from 0.3 to 5.9% within subgenus *Clerodendrum* (excluding section *Konocalyx*) and from 1.0 to 7.7% within the group comprising subgenus *Cyclonema* and section *Konocalyx*. Pairwise comparison of taxa between these two major groupings revealed approximate (due to the region of ambiguous sequence alignment between groups) sequence divergence values from 7.2 to 13.7%.

Sequence divergence values for the cpDNA were lower than those observed in the nrDNA. Approximately 4890 bp were sampled indirectly in each accession, using ten 6-bp recognition site enzymes and 43 cpDNA probes. Estimated cpDNA sequence divergence ranged from 0 to 1.8% within subgenus *Clerodendrum* (excluding section *Konocalyx*) and from 0 to 0.96% within *Cyclonema* plus section *Konocalyx*. Between these two taxonomic groupings the range was 4.3 to 5.2%.

G + C contents are shown in Table 3. The higher G + C content observed in members of Clade IV and the outgroup taxa relative to taxa in Clades I–III may account for the difficulty experienced in sequencing some outgroup species and members of subgenus *Cyclonema*, because high G + C levels can cause stronger template secondary structures, which can confound sequencing reactions.

Parsimony analysis of the nrDNA sequence data yielded 20 equally most parsimonious trees of 1018 steps (Fig. 1; Consistency Index, excluding uninformative characters, CI = 0.557; Retention Index, RI = 0.713). The CI value indicates a relatively low level of homoplasy relative to the number of taxa included (Sanderson and Donoghue, 1989), but when adjusted for taxonomic level is comparable to other sequencing studies (Givnish and

TABLE 3. Estimates of the proportion of G + C in the ITS plus 5.8S nrDNA.

Clade	Range	Average
I	0.540–0.582	0.567 ± 0.013
II	0.549–0.583	0.573 ± 0.011
III	0.569–0.594	0.586 ± 0.009
I–III	0.540–0.594	0.574 ± 0.014
IV	0.623–0.665	0.641 ± 0.015
<i>Faradaya</i> / <i>Oxera</i>	0.599–0.625	0.612 ± 0.018
<i>Caryopteris</i> / <i>Trichostema</i>	0.650–0.701	0.676 ± 0.036
<i>Karomia</i>	0.651	0.651

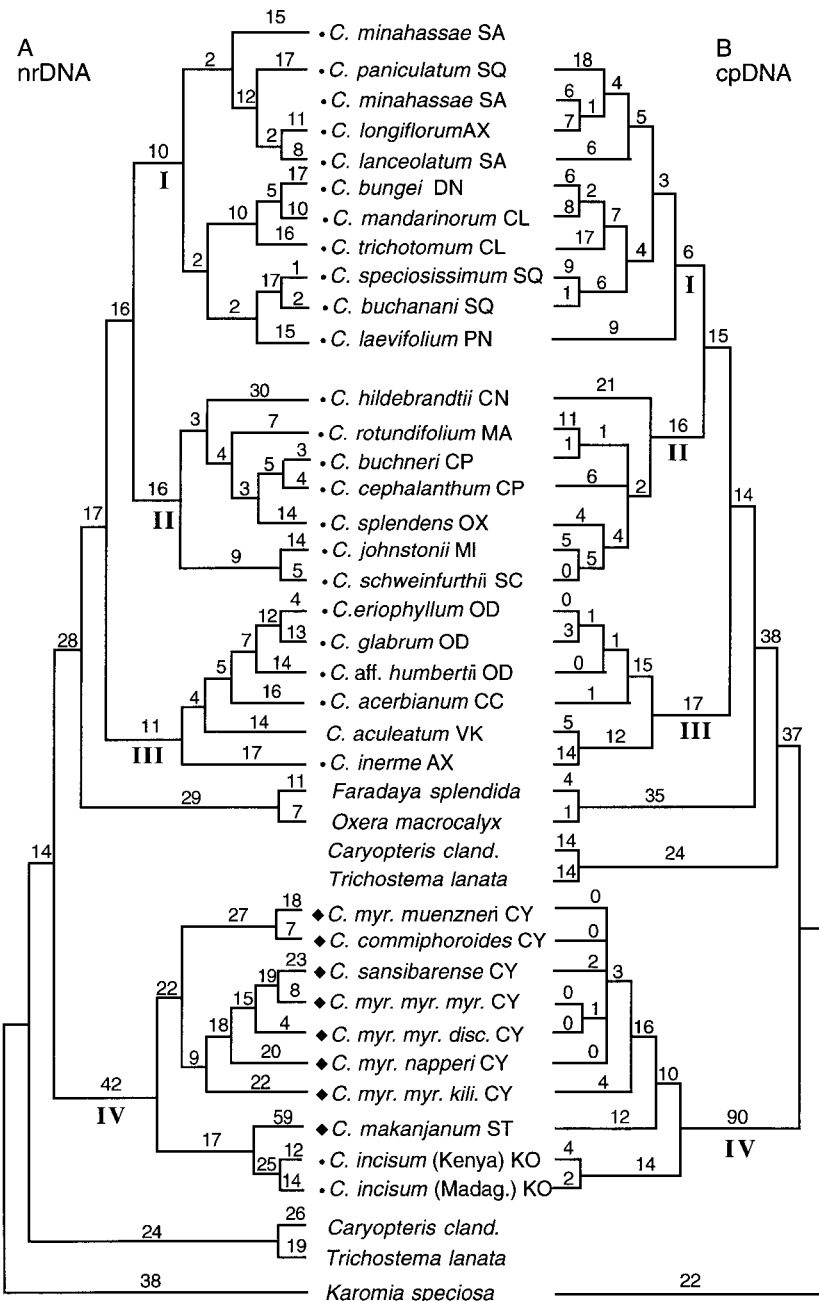


Fig. 1. (A) One of the 20 most parsimonious trees of 1018 steps derived from cladistic analysis of ITS sequence data. (B) One of the 12 most parsimonious trees of 646 steps obtained after cladistic analysis of cpDNA restriction site data. Numbers above branches represent branch lengths. Clades I–IV are indicated. *Clerodendrum* species traditionally assigned to subgenus *Clerodendrum* are marked with a dot and those belonging to subgenus *Cyclonema* are marked with a diamond. VK, Subgenus *Volkameria*. The section to which each species of *Clerodendrum* has been assigned (Moldenke, 1985; Verdcourt, 1992; see Steane et al., 1997) is indicated after the species name: AX, *Axilliflora*; CC, *Cornacchinia*; CL, *Clerodendrum*; CN, *Cylindrocalyx*; CP, *Capitata*; CY, *Cyclonema*; DN, *Densiflora*; KO, *Konocalyx*; MA, *Macrocalyx*; MI, *Microcalyx*; OD, *Odontocalyx*; OX, *Oxycalyx*; PN, *Penduliflora*; SA, *Siphonanthus*; SC, *Siphonocalyx*; SQ, *Squamata*; ST, *Stacheocymosa*. Other abbreviations: claud., *clandonensis*; disc., *discolor*; kili., *kilimandscharense*; myr., *myricoides*; Madag., Madagascar. See Appendix 1 in Steane et al. (1997) for collection details of each accession.

Sytsma, 1997). Bootstrap values of 91, 92, 96, and 95% supported Clades I, II, III, and IV, respectively (Fig. 2A). *Clerodendrum* s.s. (i.e., Clades I–III) was supported by a bootstrap value of 96%. The position of *Faradaya* and *Oxera* as the sister group to *Clerodendrum* s.s. (i.e.,

Clades I–III) is in agreement with the results obtained from the cpDNA restriction site data (Fig. 2B; Steane et al., 1997). *Caryopteris* and *Trichostema* formed the sister group to *Clerodendrum* s.l. (i.e., Clades I–IV) plus *Faradaya* and *Oxera*, rather than forming the sister group to

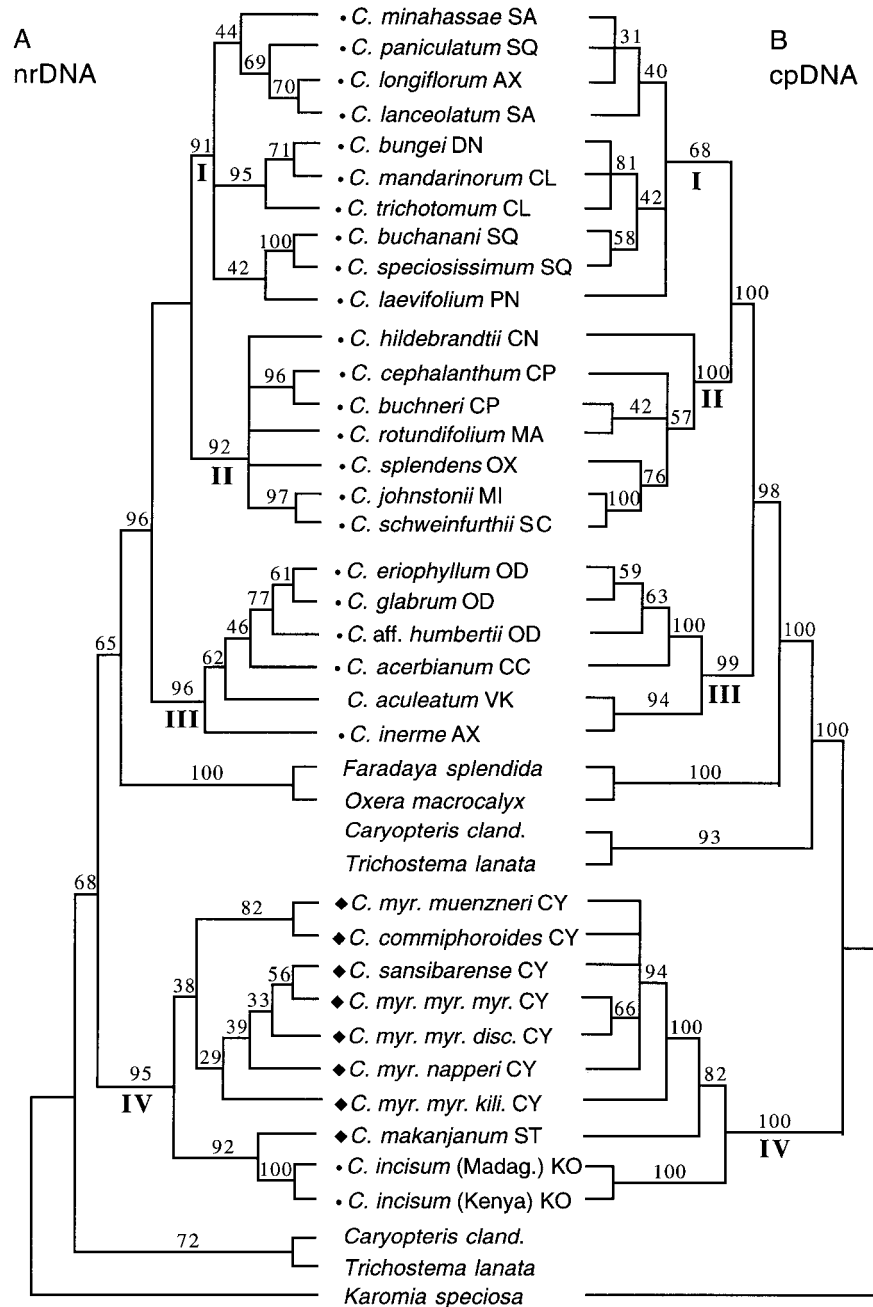


Fig. 2. Comparison of strict consensus trees derived from (A) nuclear ribosomal ITS sequence data and (B) cpDNA restriction site data. Bootstrap percentages are shown above branches. Clades I–IV are indicated. Abbreviations and symbols are as in Fig. 1.

Clerodendrum s.s. (i.e., Clades I–III) plus *Faradaya* and *Oxera*, as in the cpDNA results (Figs. 1B, 2B; Steane et al., 1997).

A subset of the taxa included in an earlier analysis of *Clerodendrum* cpDNA (Steane et al., 1997) was selected to correspond to those taxa for which sequence data were available. As a result ~5% of the 456 potentially informative characters from the original data set (Steane et al., 1997) became constant and 23% became autapomorphic. Sequence data for *Holmskioldia* were not available, so the cladograms from the smaller cpDNA analysis and from the nrDNA analysis were rooted with *Karomia* as

the outgroup (see Steane et al., 1997). Analysis of the data produced 12 equally most parsimonious trees, each 646 steps long, including autapomorphies [Fig. 1B; CI = 0.672, RI = 0.915]. A strict consensus tree was computed (Fig. 2B) that is largely congruent with the consensus tree from the complete set of taxa (Steane et al., 1997). There are a few areas of conflict between the two cladograms. In Clade I of the strict consensus of the large-scale analysis, *C. paniculatum* emerges in a clade with *C. japonicum*, *C. speciosissimum*, and *C. buechanani* (section *Squamata*). In the strict consensus tree of the smaller data set, *C. paniculatum* emerges in a clade with *C. longiflorum*

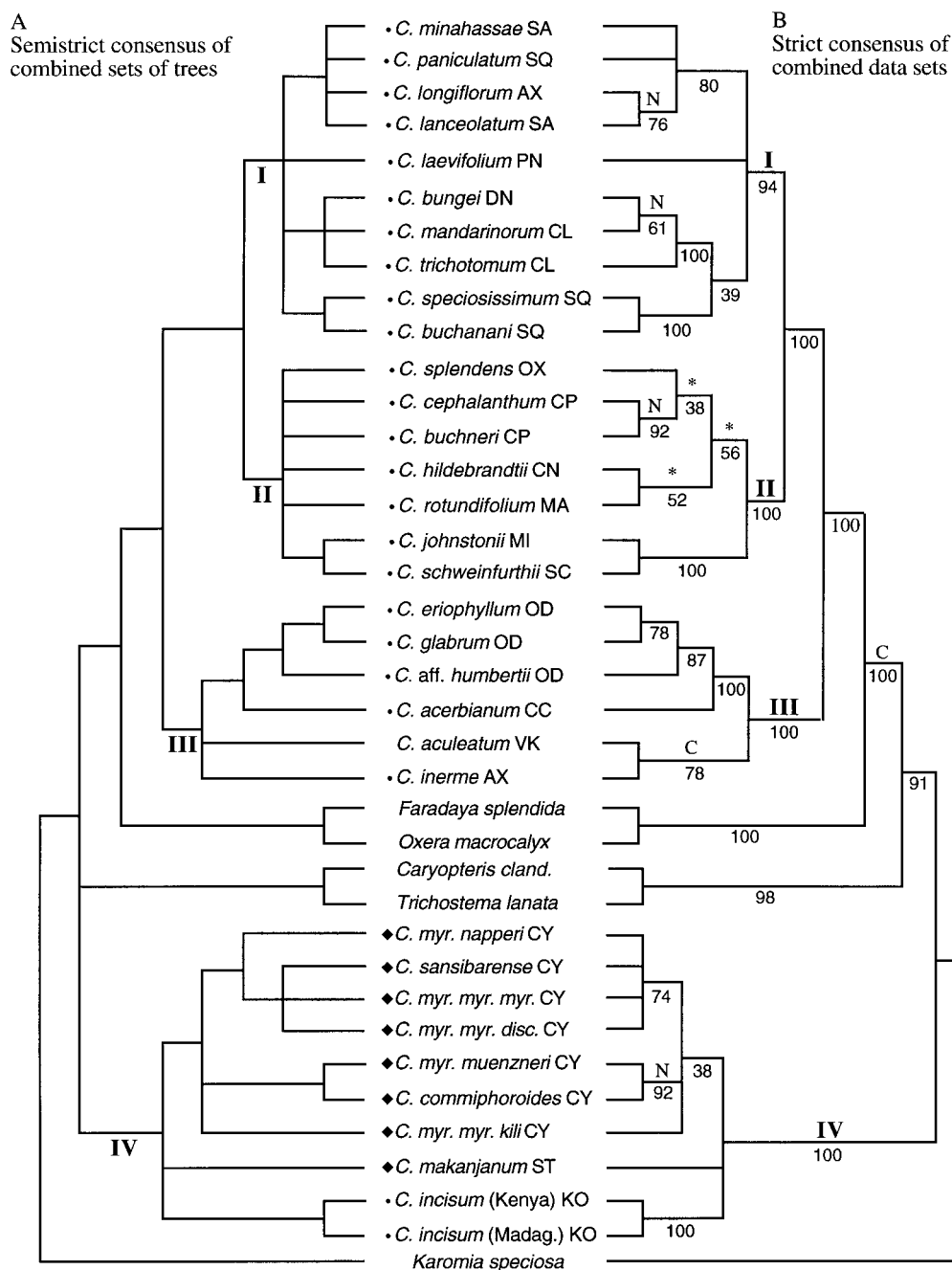


Fig. 3. (A) Semistrict consensus of 20 trees from nrDNA ITS sequence data and 12 trees from cpDNA restriction site data. (B) Strict consensus of 120 most parsimonious trees of 1687 steps obtained after cladistic analysis of combined sets of nrDNA ITS sequence data and cpDNA restriction site data. Bootstrap percentages are shown below branches. C and N, above branches, indicate clades that were resolved independently in either the cpDNA or the nrDNA analyses, respectively. Asterisks indicate clades that were not resolved by either of the independent analyses. Clades I–IV are indicated. Abbreviations and symbols are as in Fig. 1.

(section *Axilliflora*) and *C. minahassae* (section *Siphonanthus*). In Clade II of the strict consensus of the large-scale analysis, *C. hildebrandtii* and *C. rotundifolium* emerge as sister taxa. In the strict consensus of the smaller data set, *C. buchneri* emerges as the sister species to *C. rotundifolium*. Other differences between the smaller and larger analyses result from a lack of resolution rather than incongruent topologies.

Clades II and IV arise in 100% of the bootstrap replicates; Clade III is supported by a bootstrap proportion of 99% and Clade I is supported by a bootstrap proportion of 68%. *Clerodendrum* s.s. has 98% bootstrap support. *Faradaya* and *Oxera* appear as the sister group to *Clerodendrum* s.s. (Clades I–III) in a clade with 100% bootstrap support.

The set of 12 equally most parsimonious trees derived

TABLE 4. *P* values obtained from 500 replicates of the Incongruence Length Difference test.

Taxon partition	<i>P</i> ^a
All taxa	0.002 (S)
<i>Clerodendrum</i> s.l. (Clades I–IV)	0.032 (S)
<i>Clerodendrum</i> s.s. plus <i>Faradaya</i> and <i>Oxera</i>	0.092 (ns)
<i>Clerodendrum</i> s.s. (without <i>Faradaya</i> and <i>Oxera</i>)	0.146 (ns)
Clade IV	0.080 (ns)
Clade III	1.000 (ns)
Clade II	0.069 (ns)
Clade I	0.372 (ns)

^a S = significant; ns = not significant.

from the cpDNA data was combined with the set of 20 trees derived from the nrDNA sequence data to produce a set of 32 trees; there were no trees common to the two original sets. In the semistrict consensus tree (Fig. 3A), the four major clades of *Clerodendrum* (Steane et al., 1997) are preserved, with *Faradaya* and *Oxera* forming the sister group to Clades I–III. Topological incongruence within Clades I–IV between the cpDNA analysis and the nrDNA analysis are manifest in the semistrict consensus of both sets of trees by reduced resolution within Clades I–IV (Fig. 3A) and a trichotomy between (1) subgenus *Cyclonema* and section *Konocalyx* (Clade IV), (2) subgenus *Clerodendrum* (excluding section *Konocalyx*) plus *Faradaya* and *Oxera*, and (3) *Caryopteris* and *Trichostema*.

The ILD test evaluates the null hypothesis that a chosen partition is a random partition of the data. Rejection of that null hypothesis means that there may be some meaningful conflict between the data sets that cannot be attributed to sampling error. The ILD test results on the combined data set (Table 4) indicate that the cpDNA and nrDNA partitions were significantly different from random partitions of the combined data ($P = 0.01$). This suggested that caution should be exercised when combining the two data sets. The taxa were partitioned into each of the four clades (I–IV), and each partition was analyzed separately in order to determine whether the conflict in the data sets was localized in a particular region of the analysis. Significant conflict was not localized within any single clade (all had *P* values >0.05 ; Table 4), nor was there significant conflict in *Clerodendrum* s.s. (Clades I–III combined).

The sequence data made up 46% of the total informative characters (Table 1). Analysis of the combined data set produced 120 equally most parsimonious trees of 1687 steps (Fig. 4; CI = 0.581; RI = 0.841). The strict consensus tree (Fig. 3B) is well resolved in comparison to the semistrict consensus tree of the combined sets of trees (Fig. 3A). *Clerodendrum* s.l. is divided among four well-supported clades (I–IV), corresponding to those obtained from independent analyses of the two data sets (e.g., Fig. 2). Clade I is supported by a bootstrap value of 94% and Clades II, III, and IV are all found in 100% of bootstrap replicates. *Clerodendrum* s.s. (Clades I–III) has 100% bootstrap support. The clade comprising *Faradaya* and *Oxera*, the clade comprising *Trichostema* and *Caryopteris*, and Clade IV form successive sister groups to Clades I–III (*Clerodendrum* s.s.), in agreement with the results of the cpDNA-based analysis (Fig. 2B).

The strict consensus contains clades particular to one or the other or both of the separate data sets (Fig. 3B). Clade III is fully resolved and congruent with the cpDNA tree. Clade II is fully resolved, containing one clade that is congruent with the nrDNA results and three clades that were not resolved by either of the independent analyses. Results of the ILD tests (Table 4) suggest, however, that the topology within Clade II should be regarded cautiously (an ILD test with *C. splendens* removed resulted in an increased *P* value of 0.802, suggesting incongruence among the data for this species). The combined analysis identified the sister-group relationship between the clade comprising *Caryopteris* and *Trichostema* and the clade comprising Clades I–III plus *Faradaya* and *Oxera*, which was originally detected in the analyses of the cpDNA restriction site data (Fig. 2B; Steane et al., 1997), but not of the nrDNA sequence data (Figs. 1, 2).

DISCUSSION

Phylogenetic analysis of the sequence data from the 5.8S nrDNA and flanking ITS sequences suggests that *Clerodendrum* s.l. is polyphyletic, as was found by analysis of cpDNA restriction site data and sequence data (Steane et al., 1997). The four clades, I–IV, that were identified in that study are evident from the nrDNA sequence data, although the topologies within clades differ between the two data sets.

All clades that are resolved in the consensus of the separate analyses are preserved in the analysis of the combined data, except for one infraspecific clade in *C. myricoides*, which is unresolved (Fig. 3). Amalgamation of the two data sets resulted in increased resolution of the unresolved regions of the semistrict consensus of the separate data sets. In regions where the two data sets agree, the combined data yield fully resolved clades and increased bootstrap support (most importantly, the resolution of Clades I–IV and the polyphyly of *Clerodendrum* s.l.; Fig. 3). There are regions in which the independent data sets do not agree, but the signal in the combined data set either yields results congruent with those from one or the other data set (e.g., within Clades I, III, and IV, and the position of *Caryopteris* and *Trichostema* relative to the ingroup; Fig. 3), or yields results that were found by neither of the data sets (Clade II; Fig. 3). In some regions where the two data sets conflict, the resulting cladogram is unresolved. For example, the placement of *C. makanjanum* is the only case in which well-supported results in each separate analysis are incongruent (82% bootstrap support for placement with section *Cyclonema* in cpDNA and 92% bootstrap support for placement with section *Konocalyx* in nrDNA; Fig. 2).

If the assumption of character independence is violated in either of the data sets, results of the analysis of the combined data may not be a good indicator of phylogeny. The ILD test (Farris et al., 1994) did not indicate significant character incongruence, within any of the individual clades or within *Clerodendrum* s.s. (Table 4). However, for the entire data set and when the data are restricted to *Clerodendrum* s.l. (no outgroups), significant incongruence is encountered. It may be significant that Clade IV and the outgroups, particularly *Karomia*, *Caryopteris*, and *Trichostema*, have higher G + C content in their ITS

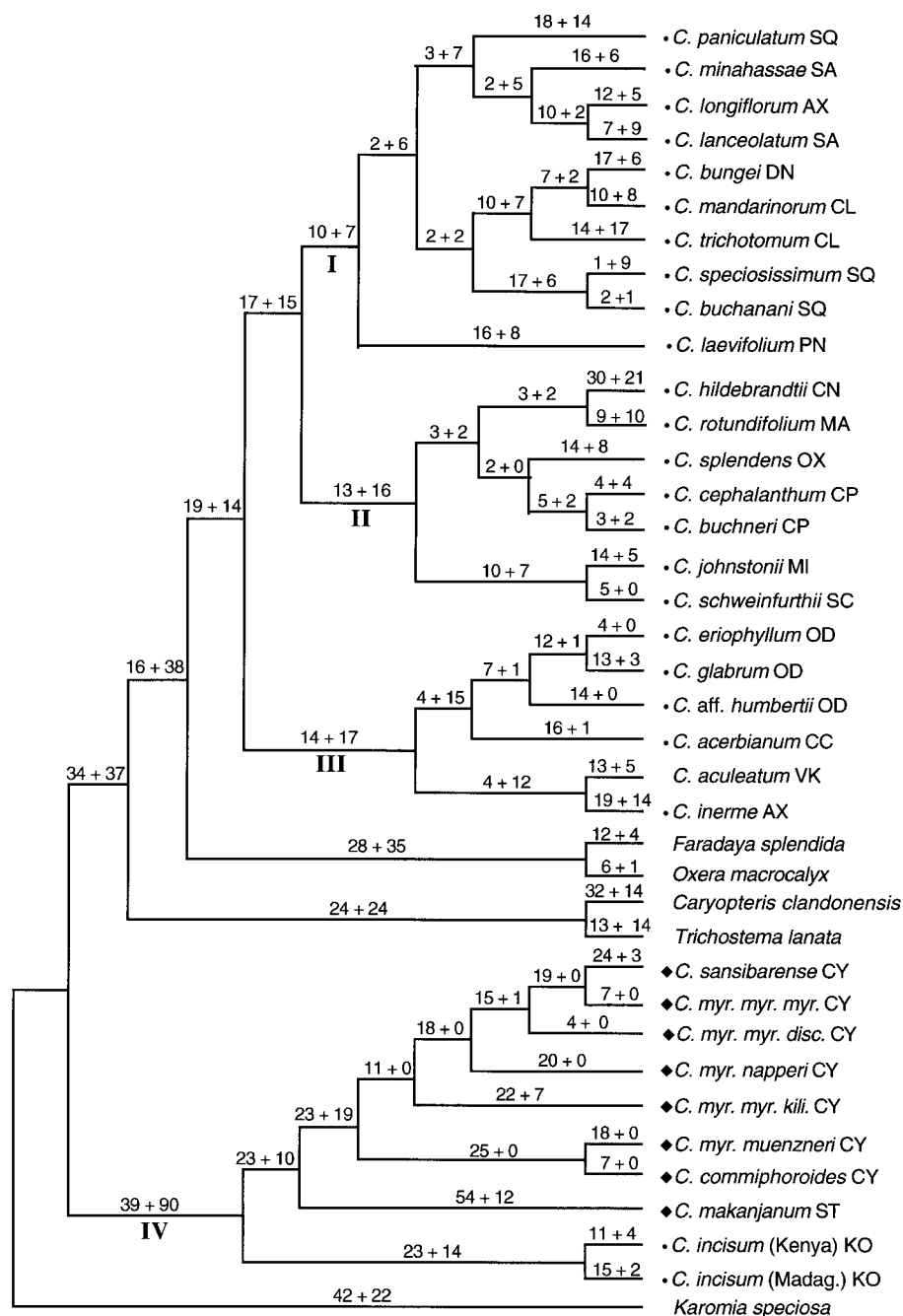


Fig. 4. One of the 120 most parsimonious trees of 1687 steps derived from combined nrDNA sequence data and cpDNA restriction site data. Branch lengths are divided into the number of base changes in the nrDNA ITS sequence data plus the number of restriction site gains/losses in the cpDNA restriction site data, respectively. Clades I–IV are indicated. Abbreviations and symbols are as in Fig. 1.

sequences than Clades I–III (Table 4). This difference in G + C content also may be responsible for the only major difference between the ITS and cpDNA trees (Fig. 2). The cpDNA data provide strong support (bootstrap = 100%) for the sister-group relationship of *Caryopteris*/*Trichostema* with *Clerodendrum* s.s. plus *Faradaya*/*Oxera*, whereas the ITS sequences weakly support (bootstrap = 68%) a basal position for *Caryopteris*/*Trichostema*. In the ITS tree (Fig. 2A), the *Caryopteris*/*Trichostema* clade, which has the highest G + C content, is clos-

est to the outgroup, *Karomia*, which has the next highest G + C content. In other, more extreme cases, differences in base composition have resulted in biased phylogenetic interpretations (Hasegawa and Hashimoto, 1993). In studies based on cpDNA restriction sites and sequences (Steane et al., 1997; Wagstaff et al., 1998), the association of *Caryopteris*/*Trichostema* with *Clerodendrum* s.s. always has been found.

The polyphyly of *Clerodendrum* s.l. is demonstrated by analysis of cpDNA restriction site data and nrDNA

data, both separately and together. All analyses agree that *Clerodendrum* s.l. is divided into four distinct clades (I–IV). The species composition of each clade is constant between analyses. Together, Clades I–III form a paraphyletic group (*ndhF* data show that *Tetradlea* is nested within *Clerodendrum* s.s.; Steane et al., 1997), comprising subgenera *Volkameria* and *Clerodendrum* except section *Konocalyx*. Section *Konocalyx* groups with subgenus *Cyclonema* in Clade IV, disjunct phylogenetically from Clades I–III. These results have important implications for the classification of *Clerodendrum* s.l. Two distinct monophyletic groups (Clades I–III plus *Tetradlea* and Clade IV) deserve recognition. Subgenus *Cyclonema* and section *Konocalyx* (subgenus *Clerodendrum* pro parte; Clade IV) should be removed from *Clerodendrum* s.l. and *Tetradlea* included (see Steane et al., 1997) for a monophyletic delimitation of *Clerodendrum* s.s. (Clades I–III). Cantino, Harley, and Wagstaff (1992) reinstated *Cyclonema* as a genus and, although they did not circumscribe the genus, Cantino (P. D. Cantino, Ohio University, personal communication) considers that section *Konocalyx* should be included in it. The results presented in this paper support the reinstatement of *Cyclonema* plus sect. *Konocalyx* as a genus (properly named *Rothea* Raf., Rafinesque, 1837; Steane and Mabberley, 1998).

The results from the molecular data also have implications for subdivision of *Clerodendrum* s.s. The genus is split into three clades of which two (Clades I and II) are definable on the basis of geography. The separation of Clade I (in which all species are Asian) and Clade II (in which all species are African) in all analyses suggests that most African and Asian taxa form genetically isolated groups. The exceptions to this rule lie in Clade III where all molecular evidence suggests that several species that inhabit coastal areas in central America (*C. aculeatum*), Asia (*C. inerme*), and Africa (*C. acerbianum*, *C. eriophyllum*, *C. glabrum*, and *C. aff. humbertii*) form a monophyletic group. The position of Clade III as sister group to Clades I and II suggests that the African and Asian species of *Clerodendrum* in the latter two clades are more closely related to each other than they are to those taxa in Clade III. The position of *Tetradlea* (from southern North America) relative to Clades I–III of *Clerodendrum* has not been established (Steane et al., 1997). The separation of Clades I, II, and III never has been acknowledged in schemes for subdividing *Clerodendrum*. Moldenke (1985) separated Asian taxa from African taxa, thus splitting Clade III, but this was probably more for convenience than through a sense of shared ancestry in each region.

Thomas (1936) and Schauer (1847) proposed systems for classifying all species of *Clerodendrum* s.l. Insufficient taxa were included in this study to determine fully the efficacy of Schauer's system of division within Clade I (Asian taxa), but the results of the large cpDNA analysis (Steane et al., 1997) suggest that Schauer's inflorescence-based system is problematic. Furthermore, to apply Thomas' scheme globally would result in the recognition of polyphyletic groups. For example, Thomas placed *C. speciosissimum* ("*C. fallax*"; sect. *Squamata* sensu Schauer) in subsection "*Obtusata*" (= subsection *Fallax* section *Oxycalyx*; Verdcourt, 1992). We disagree with this classification for two reasons: (1) all analyses of molec-

ular data separate the African and Asian *Clerodendrum* s.s. taxa (except for those coastal species in Clade III), and it is therefore unlikely that *C. speciosissimum* (Clade I) is closely related to other members of the predominantly African section *Oxycalyx* (Clade II); (2) analysis of the larger cpDNA data set (Steane et al., 1997) suggests that section *Squamata* is a monophyletic group, and this is supported by morphological and chemical characters (Stenzel et al., 1988).

Cladistic analysis of molecular data from both the chloroplast and nuclear genomes of *Clerodendrum* s.l. has demonstrated that the classification systems devised for the genus in the past (Schauer, 1847; Thomas, 1936; Moldenke 1985) inadequately reflect the natural groupings within the genus. A thorough revision of the genus is necessary. *Cyclonema* has been reinstated as a genus and delimited to include section *Konocalyx* (*Rothea* Raf.; Steane and Mabberley, 1998). The infrageneric taxa of *Clerodendrum* s.s. (Clades I–III) may be characterized by unique combinations of morphological (as well as geographical) characters, providing the foundation on which to construct a robust, functional classification that also reflects the phylogenetic history of the plants.

LITERATURE CITED

- BALDWIN, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogeny and Evolution* 1: 3–16.
- . 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA: chromosomal and morphological evolution reexamined. *American Journal of Botany* 80: 222–238.
- , M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. *Systematic Zoology* 40: 486–493.
- BRIQUET, J. 1897. Viticoideae. Clerodendreae. In A. Engler and K. Prantl [eds.], *Die Natürlichen Pflanzenfamilien IV (3a)*, 173–177. Leipzig.
- BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Botany* 42: 348–397.
- CANTINO, P. D. 1992. Evidence for a polyphyletic origin of the Labiatae. *Annals of the Missouri Botanical Garden* 79: 361–379.
- , R. M. HARLEY, AND S. J. WAGSTAFF. 1992. Genera of Labiatae: status and classification. In R. M. Harley and T. Reynolds [eds.], *Advances in labiate science*, 511–522. Royal Botanic Gardens, Kew.
- DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematics* 26: 657–681.
- DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.
- FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. *Cladistics* 10: 315–319.
- GIVNISH, T. J., AND K. J. SYTSMAN. 1997. Homoplasy in molecular vs. morphological data: the likelihood of correct phylogenetic inference. In T. J. Givnish and K. J. Sytsma [eds.], *Molecular evolution and adaptive radiation*, 55–101. Cambridge University Press, New York, NY.
- GONZALEZ, I. L., J. E. SYLVESTER, T. F. SMITH, D. STAMBOLIAN, AND R. D. SCHMICKEL. 1990. Ribosomal RNA gene sequences and hominoid phylogeny. *Molecular Biology and Evolution* 7: 203–219.
- HARRIS, S. A., AND R. INGRAM. 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon* 40: 393–412.

- HASEGAWA, M., AND T. HASHIMOTO. 1993. Ribosomal RNA trees misleading? *Nature* 361: 23.
- HIGGINS, D. G., A. J. BLEASBY, AND R. FUCHS. 1991. CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* 8: 189–191.
- HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics* 18: 23–42.
- KALTENBOECK, B., J. W. SPATAFORA, X. ZHANG, K. G. KOUSOULAS, M. BLACKWELL, AND J. STORZ. 1992. Efficient production of single-stranded DNA as long as 2 kb for sequencing of PCR-amplified DNA. *BioTechniques* 12: 164–171.
- KIM, K.-J., AND R. K. JANSEN. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, *Asteraceae*)—additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution* 190: 157–185.
- KISS, T., M. KIS, S. ABEL, AND F. SOLYMOSY. 1988. Nucleotide sequence of the 17S–25S region from tomato rDNA. *Nucleic Acids Research* 16: 7179.
- LAM, H. J. 1919. The Verbenaceae of the Malayan archipelago, together with those from the Malayan Peninsula, the Philippines, the Bismarck-archipelago, and the Palau-, Marianne- and Caroline-islands. De Waal, The Netherlands.
- LEE, S. B., AND J. W. TAYLOR. 1992. Phylogeny of five fungus-like protist *Phytophthora* species, inferred from the internal transcribed spacers of ribosomal DNA. *Molecular Biology and Evolution* 9: 636–653.
- MISHLER, B. D. 1994. Cladistic analysis of molecular and morphological data. *American Journal of Physical Anthropology* 94: 143–156.
- MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. *Cladistics* 1: 186–189.
- , AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* 44: 64–76.
- MOGENSEN, H. L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383–404.
- MOLDENKE, H. N. 1985. Notes on the genus *Clerodendrum* (Verbenaceae), IV. *Phytologia* 57: 334–365.
- OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467–481.
- PENNINGTON, R. T. 1996. Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. *Systematic Biology* 45: 456–515.
- RAFINESQUE, C. S. 1837. *Flora Telluriana*, p. 69. Rafinesque: Philadelphia, PA.
- RIESEBERG, L. H., AND D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.
- RIMPLER, H., C. WINTERHALTER, AND U. FALK. 1992. Cladistic analysis of the subfamily Caryopteridoideae Briq. and related taxa of Verbenaceae and Lamiaceae using morphological and chemical characters. In R. M. Harley and T. Reynolds [eds.], *Advances in labiate science*, 39–54. Royal Botanic Gardens, Kew.
- SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43: 1781–1795.
- SCHAUER, J. C. 1847. Verbenaceae. In A. De Candolle [ed.], *Prodromus systematis naturalis regni vegetabilis* 11, 658–676. Masson, Paris.
- STEANE, D. A., AND D. J. MABBERLEY. 1998. *Rothea* (Labiatae) revived. *Novon*, in press.
- , R. W. SCOTLAND, D. J. MABBERLEY, S. J. WAGSTAFF, P. A. REEVES, AND R. G. OLMEAD. 1997. Phylogenetic relationships of *Clerodendrum* s.l. (Lamiaceae) inferred from chloroplast DNA. *Systematic Botany* 22: 229–244.
- STENZEL, E., J. HENI, H. RIMPLER, AND D. VOGELLEHNER. 1988. Phenetic relationships in *Clerodendrum* (Verbenaceae) and some phylogenetic considerations. *Plant Systematics and Evolution* 159: 257–271.
- SUH, Y., L. B. THIEN, H. E. REEVE, AND E. A. ZIMMER. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80: 1042–1055.
- SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.1., Computer program distributed by the Illinois Natural History Survey, Champaign, IL.
- THOMAS, B. 1936. Die Gattung *Clerodendrum* in AFRIKA. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* 68: 1–106.
- VERDCOURT, B. 1992. *Clerodendrum*. In R. M. Polhill [ed.], *Flora of Tropical East Africa*, 84–144. Balkema, Rotterdam.
- WAGSTAFF, S. J., P. A. REEVES, L. HICKERSON, R. E. SPANGLER, AND R. G. OLMEAD. 1998. Phylogeny of Labiatae s.l. inferred from cpDNA sequences. *Plant Systematics and Evolution* 209: 265–274.
- WHITE, T. J., T. BRUNS, S. LEE, AND J. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. D. Innis, D. Gelfand, J. Sninsky, and T. White [eds.], *PCR protocols: a guide to methods and applications*, 315–322. Academic Press, San Diego, CA.
- WINTERHALTER, C. 1991. Cladistische UNTERSUCHUNGEN zur GLIEDERUNG der Clerodendreae und Vitaceae (Verbenaceae). Ph.D. dissertation, Albert-Ludwigs-Universität, Freiburg.
- WOJCIECHOWSKI, M. F., M. J. SANDERSON, B. G. BALDWIN, AND M. J. DONOGHUE. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 80: 711–722.
- YUAN, Y.-M., P. KÜPPER, AND J. J. DOYLE. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *American Journal of Botany* 83: 641–652.